General Biostatistics

Part 4

Methods for Statistical Inference
Hypothesis Testing

Marie Diener-West, Ph.D.
Department of Biostatistics
Johns Hopkins University
Bloomberg School of Public Health

Outline

- Rationale for hypothesis testing
- Steps in hypothesis testing
- Possible errors associated with hypothesis testing
- p-value
- Test statistic
- Examples

Rationale for Hypothesis Testing

- Aids the clinician or researcher in reaching a decision concerning a population by examining a sample
- If a hypothesis regarding the true (but unknown) population parameter is true, is the value of the sample statistic likely or unlikely?
- Researcher may reject or not reject the hypothesis on the basis of the data

Steps in Hypothesis Testing

1 Select the probability model for the observed data.

2 Set up a *null* hypothesis (H_o) and *alternative* hypothesis (H_a)
   - two-sided test
   - one-sided test

3 Select a test statistic, Z or t

Steps in Hypothesis Testing

4 Select a decision rule and critical region (rejection region); choose the significance level α

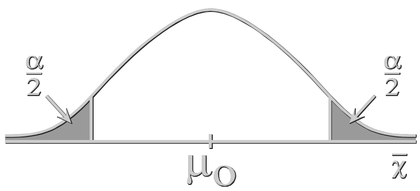
5 The critical region consists of test statistics having very low probability

6 Compute the observed value of the test statistic

7 Make a statistical decision and conclusion

Sampling Distribution

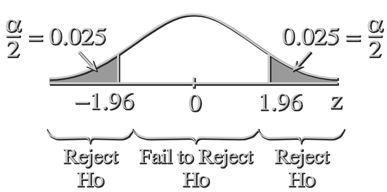
$Ho : \mu = \mu_0$



Critical Region

(using Z table)

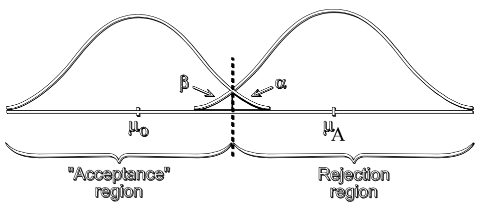
$\alpha = 0.05$



Statistical Decision

- "Reject" H_o because the value of the test statistic is very unlikely when H_o is true (values in the critical or rejection region)
- "Accept" ("Fail to reject") H_o because the value of the test statistic is likely when H_o is true (values in the "acceptance" region)

Results of Hypothesis Testing



Outcomes of Hypothesis Testing

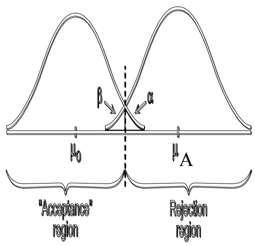
	Conclusion Based on the Data (Sample)	
	Do not reject H_o	Reject H_o
Truth: H_o is true	Correct conclusion	Type I error
Truth: H_o is false	Type II error	Correct conclusion

Possible Errors in Testing

- Type I error = reject a true Ho
- α = the probability of a Type I error
- Type II error = not reject a false Ho
- β = the probability of a Type II error
- Power = 1 - β = the probability of correctly rejecting Ho

Possible Errors in Testing

- Type I error
- Type II error
- Power of a statistical test
- Aim to keep Type I error small
- Aim to keep Type II error small and power high



p-value

- The p-value for a hypothesis test = the probability of obtaining the value of the test statistic that you obtained (or one more extreme) just by chance alone <u>when Ho is true</u>
- May be one-sided or two-sided

Interpreting a p-value

- $p < 0.05$ "statistically significant"
- $p > 0.05$ NS "not statistically significant"
- Statistical vs. Practical Significance

Test Statistic

- In general

$$\text{test statistic} = \frac{\text{(sample statistic - hypothesized value)}}{\text{standard error of the sample statistic}}$$

- Need to know
  - calculated sample statistic
  - hypothesized value of population parameter
  - standard error

Test Statistic for a Mean

- Ho: $\mu = \mu_0$
- When σ is known, use z $z = \frac{(\bar{X} - u_0)}{\frac{\sigma}{\sqrt{n}}}$

- When σ is not known, use t $t = \frac{(\bar{X} - u_0)}{\frac{s}{\sqrt{n}}}$
- When n is large, $t \Rightarrow z$

Test Statistic for a Proportion

- Ho: $p = p_0$
- Use z

$$z = \frac{(\hat{p} - p_0)}{\sqrt{\frac{p_o q_o}{n}}}$$

Example: Birth Weights

- Suppose we have taken a sample of 25 infants whose mean birth weight is 2500 gm.
- Suppose we know that full-term infants have a birth weight of 3000 gm.
- Can we conclude that we have sampled from a population of full-term infants?

Example: Birth Weights

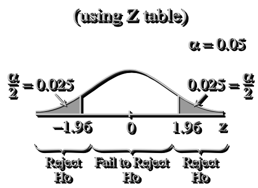
- Ho: μ = 3000
- We know n=25, $\overline{X}$ = 2500 g
- Suppose σ= 1000 gm
- Test statistic
- $$z = \frac{(\overline{X} - u_0)}{\frac{\sigma}{\sqrt{n}}} = \frac{(2500 - 3000)}{\frac{1000}{\sqrt{25}}} = -2.5$$

Example: Birth Weights

- Z=-2.5 is in the rejection region.
- Conclusion: Reject Ho that this sample is drawn from a population with a mean equal to 3000 g
- p -value = P(z < -2.5) + P(z>2.5) = 0.012

Example: Birth Weights

- p -value = 0.010
- Only 10 out of 1000 times would we observe a sample mean of 2500 gm if the true population mean is 3000 gm
- Conclusion?



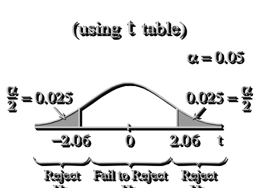
Example: Birth Weights

- Ho: μ = 3000
- Suppose s= 900g
- Test statistic:
- When n is large, t $\Rightarrow$ z

$$t = \frac{(\overline{X} - u_0)}{\frac{s}{\sqrt{n}}} = \frac{(2500 - 3000)}{\frac{900}{\sqrt{25}}} = -2.78$$

Example: Birth Weights

- p -value = 0.010
- Only 10 out of 1000 times would we observe a sample mean of 2500 g if the true population mean is 3000 g
- Conclusion?



Example: Birth Weights

- Conclusions:
  - Based on our data, it is unlikely that we sampled from a population of infants with mean birth weight of 3000 gm
  - Based on our data, it appears that we sampled from a population of infants with mean birth weight < 3000 gm
  - Our conclusion agrees with the interpretation of the 95% CI

Estimation vs. Hypothesis Testing

- The p-value is only a guideline
- Confidence interval may be more informative
- Statistical versus clinical (practical) significance
  - Related to sample size

Summary

- Hypothesis testing
  - State null hypothesis (Ho) and alternative hypothesis (Ha)
  - Use the theoretical sampling distribution to determine whether your observed sample statistic is “likely” or “unlikely” if Ho is true
  - Set critical region; calculate p-value
  - Errors associated with hypothesis testing

Summary (continued)

- Statistical decision
  - p-value calculates the probability of observing what you observed (or something more extreme) just by chance alone
    - Guidelines:
    - $p > 0.05$ If Ho is true, you are likely to observe
    - $p < 0.05$ If Ho is true, you are unlikely to observe